MBMG
Metabarcoding & Metagenomics

Metabarcoding and Metagenomics 5: 73–81
DOI 10.3897/mbmg.5.67533

Methods

Wet grinding of invertebrate bulk samples – a scalable and cost-efficient protocol for metabarcoding and metagenomics

Dominik Buchner[1], **Peter Haase**[2,3,4], **Florian Leese**[1,2]

1 *University of Duisburg-Essen, Aquatic Ecosystem Research, Universitaetsstr. 5, 45141, Essen, Germany*
2 *Centre for Water and Environmental Research (ZWU), Universitaetsstr. 3, 45141, Essen, Germany*
3 *Senckenberg Research Institute and Natural History Museum Frankfurt, Clamecystr. 12, 63571, Gelnhausen, Germany*
4 *Faculty of Biology, University of Duisburg-Essen, Universitaetsstr. 5, 45141, Essen, Germany*

Corresponding author: Dominik Buchner (dominik.buchner@uni-due.de)

Academic editor: Kat Bruce | **Received** 18 April 2021 | **Accepted** 29 June 2021 | **Published** 20 July 2021

Abstract

Most metabarcoding protocols for invertebrate bulk samples start with sample homogenisation, followed by DNA extraction, amplification of a specific marker region, and sequencing. Many of the above-mentioned laboratory steps have been verified thoroughly and best practice strategies exist, yet, no clear recommendation for the basis of almost all metabarcoding studies exists: the homogenisation of samples itself. Two different categories of devices are typically used for homogenisation: bead mills or blenders. Both have upsides and downsides. Bead mills rely on single-use plastics and therefore produce a lot of waste and are expensive. In addition to that, processing times can go up to 30 minutes making them unsuitable for large-scale studies. Blenders can handle larger sample volumes in a shorter time, and be cleaned – yet suffer from an increased risk of cross-contamination. We aimed to develop a fast, robust, cheap, and reliable sample homogenisation protocol that overcomes limitations of both approaches, i.e. does not produce difficult to discard waste and avoid single-use plastics while reducing overall costs. We tested the performance of the new protocol using six size-sorted Malaise trap samples and six unsorted stream macroinvertebrate kick-net samples. We used 14 replicates per sample and included many negative controls at different steps of the protocol to quantify the impacts of i) insufficient homogenisation and ii) cross-contamination. Our results show that 3-min homogenisation is sufficient to recover about 80% of OTUs per sample in each replicate and that a non-hazardous DIY cleaning solution provides an effective and efficient way of cleaning. The improvements of the protocol in terms of speed, ease of handling, an overall reduction of costs as well as the documented reliability and robustness make it an important candidate for sample homogenisation after sampling in particular for large-scale and regulatory metabarcoding but also metagenomics biodiversity assessments and monitoring.

Key Words

bioassessment, biodiversity soup, bulk sample, community metabarcoding, DNA isolation, LTER

Introduction

DNA metabarcoding is an efficient tool to characterize invertebrate species composition in environmental samples. Starting material can be very different and include flying insect samples, soil samples, benthos, or plankton samples (Yu et al. 2012; Hajibabaei et al. 2019b; Meyer et al. 2020). Most metabarcoding protocols start with a complete sample homogenisation, although alternative approaches are also explored (Zizka et al. 2018; Marquina et al. 2019; Nielsen et al. 2019; Martins et al. 2020). For homogenisation, samples can either be already sorted ('picked') invertebrate specimens, i.e. without debris such as plant material or sand (Carew et al. 2013; Elbrecht et al. 2017), or they can be complete environmental samples including specimens and debris (Hajibabaei et al. 2019b; Pereira-da-Conceicoa et al. 2019). Given the often high demands in terms of working hours per sample for sorting invertebrate samples with hundreds to thousands of specimens (Haase et al. 2006) there is an increasing demand

Copyright *Dominik Buchner et al.* This is an open access article distributed under the terms of the Creative Commons Attribution License (CC BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

PENSOFT.

for direct sample homogenisation without picking and sorting (Blackman et al. 2019). After sample homogenisation, DNA is extracted. Different DNA extraction protocols are routinely applied in current metabarcoding studies (e.g. Majaneva et al. 2018; Elbrecht and Steinke 2019; Hajibabaei et al. 2019a). Samples are then amplified, individually labeled, and sequenced, targeting a specific gene region (e.g. commonly cytochrome oxidase subunit I (COI) for macroinvertebrates (Elbrecht et al. 2019; Elbrecht and Leese 2017; Hajibabaei et al. 2019a) before bioinformatics processing and taxonomic assignment using search engines or tools like BLAST on reference databases such as BOLD (Ratnasingham and Hebert 2007) or NCBI GenBank (Johnson et al. 2008). The resulting taxa list is then used for biological interpretations. Given the many advantages such as taxonomic resolution, speed, and the possibility for data validation, the metabarcoding analyses of bulk or environmental DNA samples are used in hundreds of research studies, but are also increasingly considered in biomonitoring programs (Hänfling et al. 2016; Elbrecht et al. 2017; Haase et al. 2018; Hering et al. 2018; Li et al. 2019; Meyer et al. 2020; Pont et al. 2021)

Many of the above-mentioned laboratory steps have been verified thoroughly and best-practice strategies exist. For example, different extraction protocols have been analyzed (Majaneva et al. 2018; Loos and Nijland 2020), DNA polymerases have been studied for biases in metabarcoding studies (Nichols et al. 2018; Sze and Schloss 2019), the choice of markers and primers has been validated to a great extent (Elbrecht and Leese 2015; Elbrecht et al. 2017, 2019; Vamos et al. 2017), and the influence of indexing on taxon recovery has been studied (Schnell et al. 2015; Zizka et al. 2019). For all aspects, good solutions for macroinvertebrate analyses exist. However, only a few studies have so far validated the first laboratory step, i.e. sample homogenisation, despite the fact that this may have a strong effect on all downstream processes.

Sample homogenisation has been done without a drying step (Majaneva et al. 2018; Hajibabaei et al. 2019b; Pereira-da-Conceicoa et al. 2019) as well as with prior drying of the specimens (Elbrecht and Leese 2015; Elbrecht and Steinke 2019; Zizka et al. 2019). While wet-grinding is probably more time-efficient and the resulting 'slur' easier to handle with a pipette, the homogenate itself has to be dried afterward, since ethanol inhibits downstream protocol steps such as DNA extraction and PCR (Schrader et al. 2012). When drying specimens before homogenisation the dry weight can be measured more easily. However, dry tissue powder can be electrostatically charged and therefore hard to handle (Elbrecht and Steinke 2019). More importantly, cross-contamination of the highly volatile components may occur.

For sample homogenisation, different devices are used. They can be divided into two main categories: bead mills and blenders. Bead mills work by accelerating small, hard particles in a closed container like a falcon tube to break down tissue into small fragments. Blenders work with a rapidly rotating blade that slices the tissue.

While most bead mills rely on single-use plastics, blenders offer the option to be cleaned. But both methods have downsides to consider: while single-use plastics are ideal in terms of avoiding cross-contamination they produce a lot of waste and are expensive compared to the costs of other parts of the workflow. Prices may vary but go up to 15 € per sample. For sufficient homogenisation, runtime varies from 2 up to 30 mins (Elbrecht et al. 2020; Beermann et al. 2021). Also, sample volumes are low compared to blenders and mostly limited to 100 ml per sample. Bigger grinding chambers are available (e.g. IKA, Staufen im Breisgau, Germany) but increase costs even more.

Blenders on the other hand can handle large sample volumes more easily (e.g. 600 ml in Pereira-da-Conceicoa et al. (2019)) and quickly, but need to be sterilized extensively. In most studies, a combination of bleaching (5–12%) and UV-radiation (up to 30 min) was used to clean the blender (Majaneva et al. 2018; Hajibabaei et al. 2019b; Pereira-da-Conceicoa et al. 2019). While being the safest way, this is not feasible in large-scale, routine biomonitoring. Furthermore, bleach is also highly corrosive to metals and must be discarded according to strict laboratory guidelines.

Here, we aimed to develop a fast, robust, cheap, and reliable sample homogenisation protocol that overcomes the above-mentioned limitations of both methods, i.e. does not produce difficult to discard waste and avoid single-use plastics. We tested the performance of the new protocol using six sorted Malaise trap and six unsorted stream kick-net samples. We used 14 replicates per sample and included many negative controls at different steps of the protocol to quantify the impacts of insufficient homogenisation and cross-contamination.

Materials and methods

Study design

The design of this study is summarized in Fig. 1.

Sample acquisition

Two different samples types were used in this study: (i) six unsorted stream kick-net samples from a study conducted by Haase and Pilotto (2019), and (ii) six size-sorted (small < 4 mm; large > 4 mm) samples from Malaise traps that had been set up in the Rhine-Mine-Observatory for 2 weeks in July 2020. The Rhine-Main-Observatory is a Long-Term Ecological Research (LTER) site (Haase et al. 2016; Mirtl et al. 2018) east of Frankfurt, Germany (https://deims.org/9f9ba137-342d-4813-ae58-a60911c3a-bc1). All samples were preserved in 96% technical ethanol.

Tissue homogenisation and cleaning

Samples were homogenized in a common kitchen blender (Mini Blender & Blender Smoothie, Homgeek, China) at 25,000 RPM for 3 min together with the preservation liq-

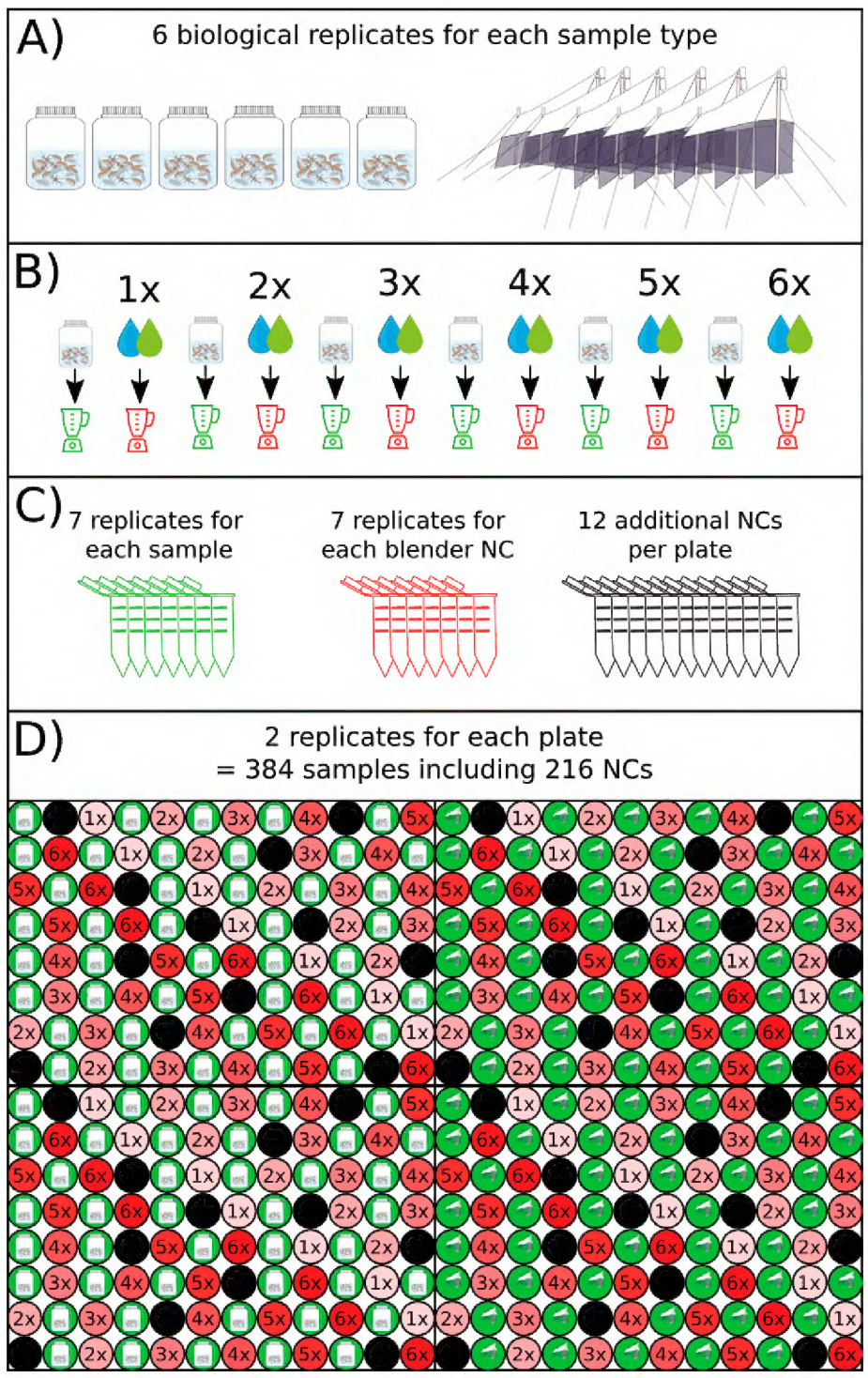

Figure 1. Schematic overview of the study design. **A)** 6 biological replicates of both sample types (stream kick-net sample, Malaise trap sample) were used in this study. **B)** Each of the samples were homogenized in the blender which was cleaned 1 – 6 times afterward by letting it run for 20 s with either ddH_2O (blue drop, kick-net samples) or self-made decontamination solution (DIY-DS, green drop, Malaise trap samples). After cleaning, the blender was filled with EtOH to create a blender negative control. **C)** Each sample, as well as each blender negative control, was replicated 7 (extraction replicates) times in 2 ml tubes. At this stage, 12 additional tubes were added that never had contact with the blender to be able to distinguish possible contamination from the sample homogenisation from contamination that occurred in the downstream analysis. **D)** The samples were then transferred into two 96-well plates, which were replicated once more (technical replicates), to distinguish between contaminations that might have happened at stage C) from contamination that might have happened after stage D).

uid. To reduce heating of the samples, samples were cooled to -20 °C prior to homogenisation. After homogenisation, samples were transferred back to their respective collection container and stored at -20 °C until DNA extraction. The blade and container of the blender were cleaned with ddH_2O until no remainders of the sample were visible. After that, the container was filled with either 100 ml of ddH_2O or self-made decontamination solution (DIY-DS, 0.6% bleach, 1% NaOH, 1% Alconox, 90 mM sodium bicarbonate, Suppl. material 1: Protocol 1), and the blender was run for 20 s to fully clean the blades and container. We here tried the DIY-DS for the first time in comparison to ddH_2O. With respect to the DIY-DS components: Low-concentration bleach is used to destroy nucleic acids and proteins. Sodium bicarbonate makes the solution less corrosive to metal surfaces (see US patent US8765652B2) while NaOH maintains a high pH high to fully denature proteins. Alconox is used as a detergent and wetting agent to effectively remove residual tissue parts. After that, container and blades were rinsed with ddH_2O and used for the next sample. This cleaning procedure was repeated 1 – 6 times to assess how many rounds of cleaning were necessary for full decontamination. Before each sample, the blender was filled with 100 ml 96% EtOH, which was treated as if it was a sample to produce a blender negative control. While all kick-net samples were treated with ddH_2O, DIY-DS was used for the Malaise trap samples. In between the homogenisation of the 2 size fractions for the Malaise traps we only shortly rinsed the blender container with ddH_2O. Twelve additional negative controls were added to the samples after the homogenisation step to control for any kind of cross-contamination independent of the blending procedure. Homogenisation of samples to microscopic tissue pieces (homogenate becomes powder upon ethanol evaporation) was optically verified using a digital microscope (Keyence VHX-6000, Keyence, Osaka, Japan). Details on the blender container can be provided on request.

DNA extraction

Before tissue lysis, the two size fractions of the Malaise trap samples were pooled in a 1:5 ratio (large-small, 5 ml and 25 ml) as suggested by Elbrecht et al. (2020). 500 µl of each sample was used for DNA extraction, which corresponds to roughly 0.1% of the complete 500 ml sample volume. All samples were centrifuged at maximum speed (14,000 × g) for 5 minutes to pellet the tissue. All remaining Ethanol was then evaporated in an Eppendorf Concentrator (Eppendorf, Hamburg, Germany) with a speed of 1400 RPM and at a temperature of 60 °C. Subsequently, 300 µl of TNES buffer was added to each tube as well as a small amount of silicon-carbide beads (0.1 mm diameter, Biospec Products, Bartlesville, USA). The tissue pellet was broken up by 30 s of bead-beating in a FastPrep Bead Beater (MP Biomedicals, Eschwege, Germany).

All subsequent processing steps were completed on a Biomek FX^P liquid handling workstation (Beckman Coulter, Brea, CA, USA). 60 µl of tissue dissolved in TNES buffer was taken out twice of every tube and mixed with 133 µl TNES and 7 µl Proteinase K (10 mg/ml) and digested for 3 h at 55 °C. From this point onwards, the plates containing replicate samples (see Fig. 1 for an overview of the replication) were never opened at the same time. DNA was extracted using a modified version (see Buchner et al. 2021) of the NucleoMag Tissue kit (Macherey Nagel, Düren, Germany). Extraction success was visualised on a 1% agarose gel.

qPCR validation, PCR, and library preparation

To control for possible contamination of the negative controls, all samples were amplified in a quantitative PCR (qPCR) in 20 µl reactions containing 1× perfeCTa FastMix, 300 nM of each primer (fwh2F, fwhR2n (Vamos et al. 2017)) and 1 µl of extracted DNA, filled up with PCR-grade water with the following cycling conditions: 30 s of initial denaturation, 80 cycles of 5 s denaturation at 95 °C, 30 s of annealing at 58 °C and 20 s of extension at 72 °C following a standard melting curve on a StepOne Plus thermocycler (Applied Biosystems, Foster City, CA, USA). The success of the qPCR was verified via melting curve analysis and visualisation on a 1% agarose gel.

The PCR for the metabarcoding library was done in a two PCR step protocol (Zizka et al. 2019). Samples were amplified in a first PCR with the Qiagen Multiplex Plus Kit (Qiagen, Hilden, Germany) with a final concentration of 1× Multiplex Mastermix, 1× Corralload Loading Dye, 100 nM of each primer (fwh2F, fwhR2n (Vamos et al. 2017)), 2.5 µl of DNA filled up to a final volume of 25 µl with PCR-grade water. For amplification, a touchdown protocol was used: 5 min of initial denaturation, 10 cycles of 30 s denaturation at 95 °C, 30 s of annealing at 68–59 °C decreased by 1 degree in each cycle, 30 s of extension at 72 °C followed by 20 cycles of 30 s denaturation at 95 °C, 30 s of annealing at 58 °C, 30 s of extension at 72 °C finished by a final elongation step of 10 min at 68 °C. For subsequent demultiplexing, each of the PCR plates was tagged with a unique combination of inline-tags. The primers also contain a universal binding site for the second step PCR primer to anneal (see Suppl. material 2: Table S1 for detailed information on the primers used in this study).

In the second PCR, samples were amplified with the Qiagen Multiplex Plus Kit with the same final concentrations except that 1 µl of first step PCR product was used as a template. For amplification the following protocol was used: initial denaturation for 5 min at 95 °C, 25 cycles of 30 s denaturation, and 60 s of combined annealing and extension at 72 °C finished with a final elongation for 10 min at 68 °C. In the second PCR, each of the 96 wells was individually tagged so that the combination of inline-tag from the first PCR step and index-read of the second step yields a unique combination. The success of the PCR was visualisation on a 1% agarose gel.

PCR product concentrations were normalised using the SequalPrep Normalisation plate (Invitrogen, Carlsbad, CA, USA). Normalised products were then pooled to the final library in equal parts for all samples. The library was concentrated using the NucleoSpin kit (Macherey Nagel, Düren, Germany) and dual-sided size selected (right ratio: 0.6; left ratio: 0.75) with the NucleoMag size-select kit (Macherey Nagel, Düren, Germany). Library concentration was quantified on a Fragment Analyzer (High Sensitivity NGS Fragment Analysis Kit; Advanced Analytical, Ankeny, USA). The library was then sequenced using the HiSeq X platform with a paired-end (2×151 bp) kit at Macrogen Europe.

Bioinformatics, filtering, and statistical analysis

For analysis of the qPCR results, raw fluorescence values were exported from the instrument and baseline-corrected with the LinRegPCR software (Ruijter et al. 2009). The baseline-corrected data was then used for inspection of amplification- and melting curves.

Raw data of the sequencing run yielded 638,892,616 reads and was delivered demultiplexed by index-reads. Index jump (sensu Schnell et al. 2015) rate for this sequencing run was estimated with a custom python script by counting all index combinations used in the study that were found in the unassigned reads and dividing the number by the sum of all reads. Further demultiplexing by inline-tag was done with the python script "demultiplexer" (v1.0.5; https://github.com/DominikBuchner/demultiplexer). Sequences were subsequently processed with the JAMP-pipeline (v0.67; https://github.com/VascoElbrecht/JAMP). Paired-end reads were merged using Usearch (v11.0.667, Edgar 2010) via the command U_merge(fastq_pctid=75). Primers were trimmed using Cutadapt (v2.5, Martin 2011), and only reads with a length of 205 bp (± 10) were retained for further analysis. Before OTU clustering with a similarity threshold of 97%, reads were dereplicated and singletons were excluded. Only clusters with at least 0.01% abundance in one sample were used in further analysis. Taxonomic assignment was carried out using BOLDigger (v1.2.2, https://github.com/DominikBuchner/BOLDigger; Buchner and Leese 2020). The best hit was determined with the BOLDigger method. This resulted in a raw read table including taxonomic assignment (Suppl. material 3: Table S2).

To control for contamination on the robotic deck, technical replicates of the plates were merged retaining only the mean read number of both replicates if reads were present in both of the replicates. After that, the maximum number of reads for each OTU in all additional negative controls was calculated and subtracted from all reads of the respective OTU to remove noise introduced by the laboratory workflows resulting in a cleaned read table (Suppl. material 4: Table S3). All cleaning steps described above were done with a custom python script (Suppl. material 5: Script 1). All figures were created using the python package "seaborn" (Waskom et al. 2020) with custom python scripts.

To compute the similarity between samples the Jaccard index was used (Jaccard 1912). Rarefaction analysis was computed with a custom python script. Samples were drawn randomly without replacement 1000 times to generate a distribution from which the plots were generated.

Results

Optical verification of homogenisation success

No identifiable parts of the animals were left after 3 min of homogenisation. However, particle size was, overall, coarser for the Malaise trap samples (Suppl. material 6: Fig. S1).

Amplification and melting curve analysis

All invertebrate samples were amplified successfully during qPCR analysis at first try in both technical replicates (Suppl. material 7: Fig. S2). Some of the blender negative controls, as well as the additional negative controls, showed a signal in later cycles that were not consistent across replicates and most likely reflected primer dimers in qPCR. The melting curve analysis showed a clear peak at above 75 °C for all of the invertebrate samples, while negative and mixing controls showed either a double peak or a single peak below 70 °C indicating the formation of primer dimers. Primer dimers were clearly visible in the agarose gel verifying this assumption. The primer used here is known to generate primer dimers (see Vamos 2017, Suppl. material 4: Table S3).

Quantification of tag jumps

Sequencing yielded 110,641,213 reads that could not be assigned to any of the index combinations used in this study. Only 15 of these reads had a combination of the used twin-indices resulting in a very low index jumping rate of 2×10^{-8}.

OTU richness and potential cross-contamination

Mean OTU richness across all seven replicates ranged from 48.71 (45–50) to 74.00 (67–83) for the kick-net samples. For the blender negative controls that were rinsed with ddH_2O, the mean richness ranged from 0.14 (0–1) to 1.71 (0–6). Regarding the blender negative controls, none of the OTUs was found in all 7 replicates. For the Malaise trap samples mean OTU richness ranged from 293.57 (279–301) to 446.71 (437–456). For the blender negative controls rinsed with DIY-DS, the mean richness ranged from 1.14 (0–4) to 4.00 (0–20). None of the OTUs was found in all 7 replicates either (Fig. 2, top panel).

The mean number of reads per sample was overall higher for the kick-net samples than for the Malaise trap samples (1.28×10^6–1.72×10^6 vs. 7.31×10^5–1.04×10^6). The mean number of reads in the blender negative controls was overall lower for the DIY-DS treatment than for the ddH_2O treatment (192 vs. 20,718). The mean read numbers were largely influenced by one OTU having a high number of reads for only one of the mixing negative controls (Fig. 2, lower panel). Out of all 96,791,297 reads in the final, cleaned OTU table, 33,281,111 reads were assigned to the Malaise trap samples and 8,053 (0.02%) were assigned to OTUs in the corresponding blender negative controls, while 61,866,043 reads were assigned to the kick-net samples while 870,196 (1.41%) were found in the blender negative controls of the kick-net samples.

Replicate consistency

Mean Jaccard similarity between the 7 extraction replicates was overall high for both sample types (kick-net samples: 0.81 vs. Malaise trap samples: 0.84) with the spread being higher for the kick-net samples (0.58 – 0.94) in comparison to the Malaise trap samples (0.78 – 0.9) mainly due to sample number 5 of the kick-net samples (Fig. 3).

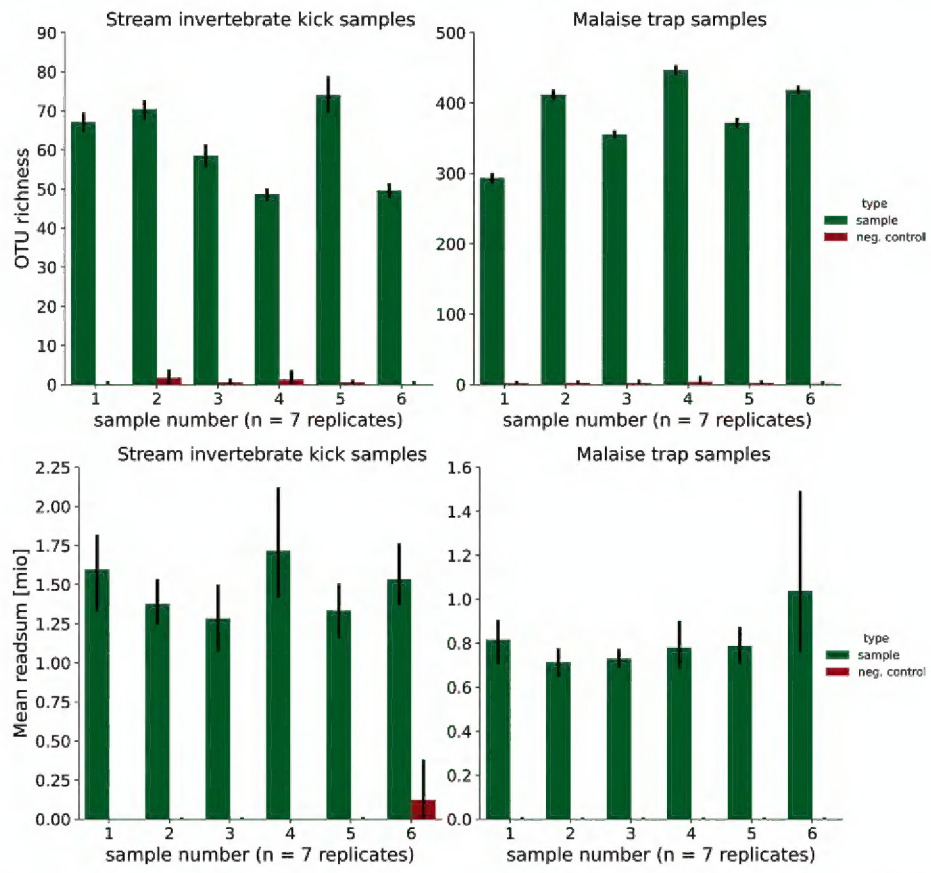

Figure 2. Mean OTU richness for kick-net samples (top left panel) and Malaise trap samples (top right panel). Mean sum of all reads across all 7 replicates for one sample for the kick-net samples (bottom left panel) and the Malaise trap samples (bottom right panel). The sample number also indicated the rounds of cleaning after each sample. Error bars indicate the 95% confidence interval ranging from percentile 2.5 to 97.5.

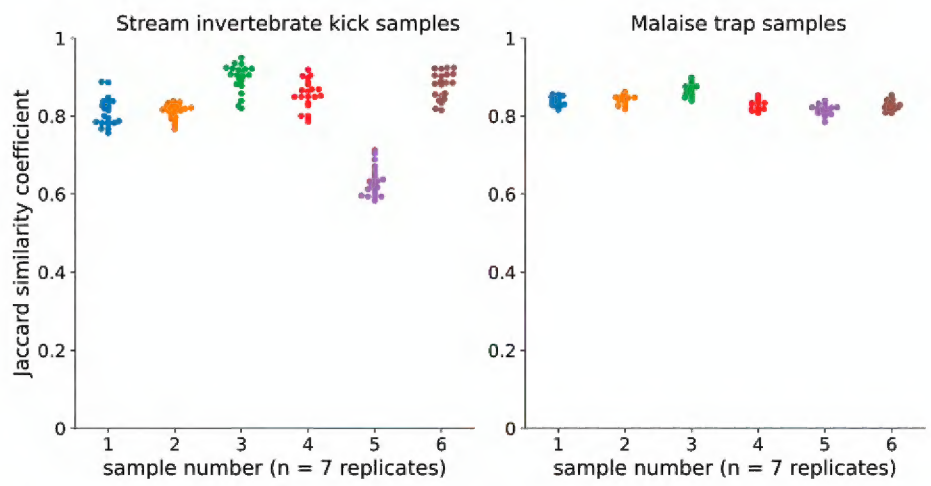

Figure 3. Pairwise comparison of extraction replicates. For each pair of extraction replicates within one sample the Jaccard similarity was computed for the kick-net samples (left panel) and the Malaise trap samples (right panel).

Rarefaction analysis for the extraction replicates of samples (Fig. 1, C) showed that one sample is sufficient in most cases to recover about 80% of the maximum possible OTU richness across all 7 extraction replicates. To recover 90% of the maximum possible OTUs, more replication effort is needed, typically >2. Of the kick-net samples, only sample number 5 is an outlier to that general pattern (Fig. 4).

Discussion

Our study aimed to develop and test and improved invertebrate homogenisation method that is easy to apply, robust,

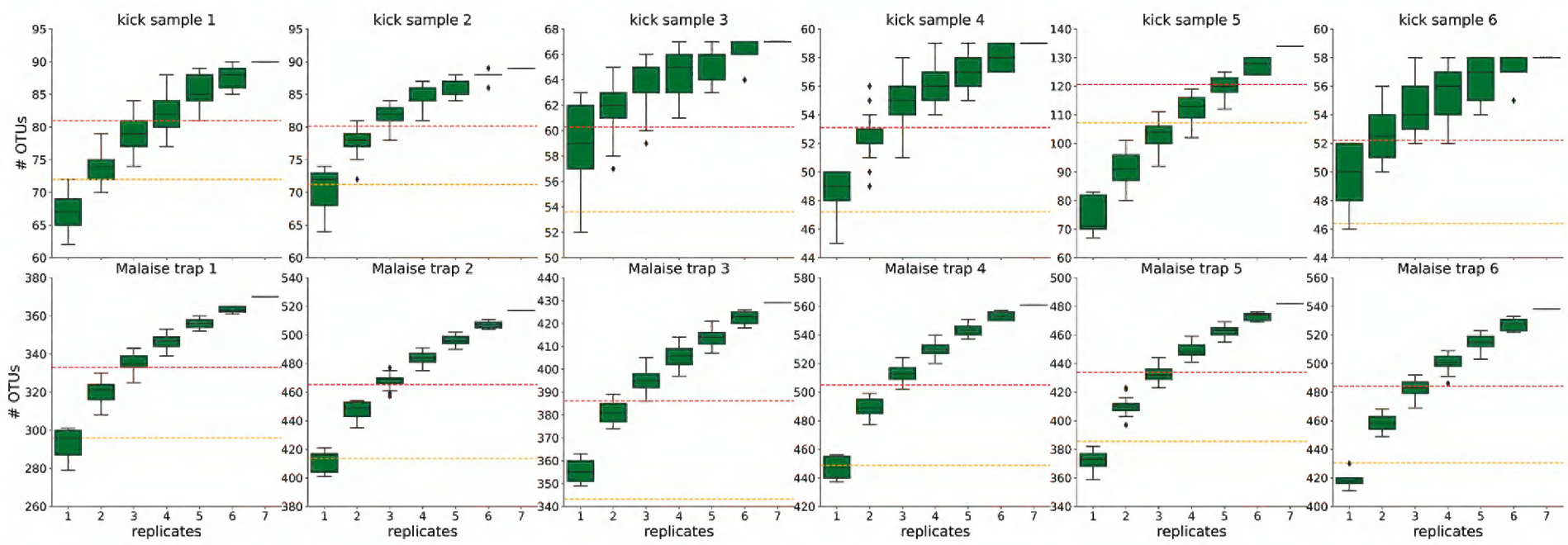

Figure 4. Rarefaction analysis of the technical replicates of each sample (**upper** row: kick-net samples, **lower row:** Malaise trap samples). Samples were randomly drawn 1000 times without replacement to generate the distribution. The yellow dashed line indicates 80% of the maximum possible value, the red dashed line indicates 90% of the maximum possible value.

and reliable, while being cost- and time efficient. Blenders are already used for species homogenisation in metabarcoding studies, however, none of those met the above mentioned criteria. Majaneva et al. (2018) used 5% bleach for 15 minutes and 30 minutes of UV sterilization, Hajibabaei et al. (2019b) used ELIMINase and also UV treatment for 30 minutes, Pereira-da-Conceicoa et al. (2019) only state that they used 12% bleach without a volume or a time. All methods produce hazardous waste or are high in costs and take at least 30 minutes per sample. In addition to that, all studies did not report OTU/ESV numbers in the negative controls or only used negative controls in PCR, which does not validate the effectiveness of the cleaning approach. In comparison, our newly developed protocol reduces sample processing time significantly without the risk of cross-contamination, which is important for the further implementation of DNA metabarcoding and metagenomic techniques in the increasingly important field of invertebrate bioassessment and monitoring.

Cross-contamination

A central concern for bulk sample metabarcoding using blenders rather than single-use plastics is the risk of cross-contamination. The approach we present minimizes this risk effectively due to three points: i) pipetting homogenized samples in ethanol (wet grinding) limits the risk of electrostatic charge and thereby 'jumping' specimens. ii) Both tested cleaning procedures, i.e. cleaning with ddH20 and the DIY-DS, proved to be highly effective. While we sporadically saw that some blender negative controls contained low read numbers of single OTUs, this was never the case for all 7 extraction replicates of the given sample. This suggests, that the contamination did not happen in the blender (or only by one of the few left molecules). This was further confirmed by the observation that some of the OTUs found in the blender negative controls were not found in the sample processed before. Furthermore, the DIY-DS reduced this already sporadic and low contamination even further and is thus recommendable. iii) The stringent replication scheme, i.e. performing extraction and downstream analysis twice in physically independent plates that are never open on the benchtop at the same time, further limits the possibility for cross-contamination. This allows to control for low-level cross-contamination by accepting reads or OTUs / ESVs that are only found in both replicates.

Replicate consistency

For the two sample types analyzed here, stream benthic macroinvertebrates and insect Malaise trap samples, we observed a high consistency between extraction replicates with typically 80% or higher OTU overlap among replicates (Jaccard similarity). Stream invertebrate kick-net sample 5 was an outlier sample with only 60–70% overlap. The reasons for the lower overlap can be insufficient blending of the sample or independent replicate contamination. Independent contamination seems unlikely, as this was not observed in any of the blender negative controls performed between samples. Microscopic inspection of homogenized tissue did also not indicate systematic differences between sample 5 and all others (Suppl. material 6: Fig. S1). However, this cannot be fully ruled out.

To further improve replicate consistency, i.e. maximize the overlap between replicates, it might be beneficial to first perform lysis on a large fraction of the sample and then perform downstream analysis using two (or more) replicates. Alternatively, and in particular, when the aim is to recover the maximum of species diversity, more extraction replicates should be performed. However, our analysis shows that we already detected 80% of the OTUs found in all 7 extraction replicates with a single replicate.

Proposition for routine application

Invertebrate assessment and monitoring using bulk DNA metabarcoding or metagenomics require fast, reliable, and validated protocols that are ideally economically competitive and environmentally friendly (Blackman et al. 2019) and yield comprehensive data on community composition. For this purpose, homogenizing samples

with blenders rather than single-use plastics seems more suitable, yet so far suffered from time-consuming and hazardous chemicals used for cleaning (highly concentrated bleach). Our study proves that the cleaning procedure can be done fast and reliably without aggressive chemicals using DIY-DS. Furthermore, the solution costs only about 1 Cent per cleaning (100 ml) making it economically feasible at the largest scale. Furthermore, while commonly used plastic or glass blenders can break when stones or hard-shelled specimens are homogenized in the blender, we here propose a solution using a stainless steel blending container (Suppl. material 8: Fig. S3).

With this container, wet grinding using ethanol preserved samples can be done fast and reliable even with large volumes of 500 ml. We could process about 30 complete bulk samples per 8 h with one person and one blender. Thus, the approach based on wet-sample grinding is not only technically feasible, scientifically reliable, economically competitive, and environmentally friendly, it offers great speed and is scalable allowing for large-scale DNA-based biomonitoring.

Acknowledgments

We thank the Leese lab members, especially Martina Weiss and Arne J. Beermann for comments and feedback on the study design. We thank Alexander Klenov (www.pipettejockey.com) for valuable input on the DIY-DS. FL is a member of and supported by COST Action DN-Aqua-Net (CA15219). This project was conducted as part of DFG project LE 2323/9-1. PH received funding from the eLTER PLUS project (Grand Agreement No. 871128).

Data accessibility statement

Demultiplexed raw read data for this publication has been uploaded to Zenodo.org and can be accessed via 10.5281/zenodo.5039930.

References

Beermann AJ, Werner M-T, Elbrecht V, Zizka VMA, Leese F (2021) DNA metabarcoding improves the detection of multiple stressor responses of stream invertebrates to increased salinity, fine sediment deposition and reduced flow velocity. Science of The Total Environment 750: e141969. https://doi.org/10.1016/j.scitotenv.2020.141969

Blackman RC, Mächler E, Altermatt F, Arnold A, Beja P, Boets P, Egeter B, Elbrecht V, Filipe AF, Jones JI, Macher J, Majaneva M, Martins FMS, Múrria C, Meissner K, Pawlowski J, Yáñez PLS, Zizka VMA, Leese F, Price B, Deiner K (2019) Advancing the use of molecular methods for routine freshwater macroinvertebrate biomonitoring – the need for calibration experiments. Metabarcoding and Metagenomics 3: e34735. https://doi.org/10.3897/mbmg.3.34735

Buchner D, Leese F (2020) BOLDigger – a Python package to identify and organise sequences with the Barcode of Life Data systems. Metabarcoding and Metagenomics 4: e53535. https://doi.org/10.3897/mbmg.4.53535

Buchner D, Beermann AJ, Leese F, Weiss M (2021) Cooking small and large portions of "biodiversity-soup": Miniaturized DNA metabarcoding PCRs perform as good as large-volume PCRs. Ecology and Evolution. https://doi.org/10.1002/ece3.7753

Carew ME, Pettigrove VJ, Metzeling L, Hoffmann AA (2013) Environmental monitoring using next generation sequencing: rapid identification of macroinvertebrate bioindicator species. Frontiers in Zoology 10: e45. https://doi.org/10.1186/1742-9994-10-45

Edgar RC (2010) Search and clustering orders of magnitude faster than BLAST. Bioinformatics 26: 2460–2461. https://doi.org/10.1093/bioinformatics/btq461

Elbrecht V, Leese F (2015) Can DNA-Based Ecosystem Assessments Quantify Species Abundance? Testing Primer Bias and Biomass–Sequence Relationships with an Innovative Metabarcoding Protocol. PLoS ONE 10: e0130324. https://doi.org/10.1371/journal.pone.0130324

Elbrecht V, Leese F (2017) Validation and Development of COI Metabarcoding Primers for Freshwater Macroinvertebrate Bioassessment. Frontiers in Environmental Science 5. https://doi.org/10.3389/fenvs.2017.00011

Elbrecht V, Steinke D (2019) Scaling up DNA metabarcoding for freshwater macrozoobenthos monitoring. Freshwater Biology 64: 380–387. https://doi.org/10.1111/fwb.13220

Elbrecht V, Vamos EE, Meissner K, Aroviita J, Leese F (2017) Assessing strengths and weaknesses of DNA metabarcoding-based macroinvertebrate identification for routine stream monitoring. Methods in Ecology and Evolution 8: 1265–1275. https://doi.org/10.1111/2041-210X.12789

Elbrecht V, Bourlat SJ, Hörren T, Lindner A, Mordente A, Noll NW, Sorg M, Zizka VMA (2020) Pooling size sorted malaise trap fractions to maximise taxon recovery with metabarcoding. bioRxiv: 2020.06.09.118950. https://doi.org/10.1101/2020.06.09.118950

Elbrecht V, Braukmann TWA, Ivanova NV, Prosser SWJ, Hajibabaei M, Wright M, Zakharov EV, Hebert PDN, Steinke D (2019) Validation of COI metabarcoding primers for terrestrial arthropods. PeerJ 7: e7745. https://doi.org/10.7717/peerj.7745

Haase P, Pilotto F (2019) A method for the reintroduction of entire benthic invertebrate communities in formerly degraded streams. Limnologica 77: e125689. https://doi.org/10.1016/j.limno.2019.125689

Haase P, Frenzel M, Klotz S, Musche M, Stoll S (2016) The long-term ecological research (LTER) network: Relevance, current status, future perspective and examples from marine, freshwater and terrestrial long-term observation. Ecological Indicators 65: 1–3. https://doi.org/10.1016/j.ecolind.2016.01.040

Haase P, Murray-Bligh J, Lohse S, Pauls S, Sundermann A, Gunn R, Clarke R (2006) Assessing the impact of errors in sorting and identifying macroinvertebrate samples. In: Furse MT, Hering D, Brabec K, Buffagni A, Sandin L, Verdonschot PFM (Eds) The Ecological Status of European Rivers: Evaluation and Intercalibration of Assessment Methods. Developments in Hydrobiology. Springer Netherlands, Dordrecht, 505–521. https://doi.org/10.1007/978-1-4020-5493-8_34

Haase P, Tonkin JD, Stoll S, Burkhard B, Frenzel M, Geijzendorffer IR, Häuser C, Klotz S, Kühn I, McDowell WH, Mirtl M, Müller F, Musche M, Penner J, Zacharias S, Schmeller DS (2018) The next generation of site-based long-term ecological monitoring: Linking essential biodiversity variables and ecosystem integrity. Science of The Total Environment 613–614: 1376–1384. https://doi.org/10.1016/j.scitotenv.2017.08.111

Hajibabaei M, Porter TM, Wright M, Rudar J (2019a) COI metabarcoding primer choice affects richness and recovery of indicator taxa in freshwater systems. PLOS ONE 14: e0220953. https://doi.org/10.1371/journal.pone.0220953

Hajibabaei M, Porter TM, Robinson CV, Baird DJ, Shokralla S, Wright MTG (2019b) Watered-down biodiversity? A comparison of metabarcoding results from DNA extracted from matched water and bulk tissue biomonitoring samples. PLOS ONE 14: e0225409. https://doi.org/10.1371/journal.pone.0225409

Hänfling B, Handley LL, Read DS, Hahn C, Li J, Nichols P, Blackman RC, Oliver A, Winfield IJ (2016) Environmental DNA metabarcoding of lake fish communities reflects long-term data from established survey methods. Molecular Ecology 25: 3101–3119. https://doi.org/10.1111/mec.13660

Hering D, Borja A, Jones JI, Pont D, Boets P, Bouchez A, Bruce K, Drakare S, Hänfling B, Kahlert M, Leese F, Meissner K, Mergen P, Reyjol Y, Segurado P, Vogler A, Kelly M (2018) Implementation options for DNA-based identification into ecological status assessment under the European Water Framework Directive. Water Research 138: 192–205. https://doi.org/10.1016/j.watres.2018.03.003

Jaccard P (1912) The Distribution of the Flora in the Alpine Zone.1. New Phytologist 11: 37–50. https://doi.org/10.1111/j.1469-8137.1912.tb05611.x

Johnson M, Zaretskaya I, Raytselis Y, Merezhuk Y, McGinnis S, Madden TL (2008) NCBI BLAST: a better web interface. Nucleic Acids Research 36: W5–W9. https://doi.org/10.1093/nar/gkn201

Li J, Hatton-Ellis TW, Handley L-JL, Kimbell HS, Benucci M, Peirson G, Hänfling B (2019) Ground-truthing of a fish-based environmental DNA metabarcoding method for assessing the quality of lakes. Journal of Applied Ecology 56: 1232–1244. https://doi.org/10.1111/1365-2664.13352

Loos LM van der, Nijland R (2020) Biases in bulk: DNA metabarcoding of marine communities and the methodology involved. Molecular Ecology n/a. https://doi.org/10.1111/mec.15592

Majaneva M, Diserud OH, Eagle SHC, Hajibabaei M, Ekrem T (2018) Choice of DNA extraction method affects DNA metabarcoding of unsorted invertebrate bulk samples. Metabarcoding and Metagenomics 2: e26664. https://doi.org/10.3897/mbmg.2.26664

Marquina D, Esparza-Salas R, Roslin T, Ronquist F (2019) Establishing arthropod community composition using metabarcoding: Surprising inconsistencies between soil samples and preservative ethanol and homogenate from Malaise trap catches. Molecular Ecology Resources 19: 1516–1530. https://doi.org/10.1111/1755-0998.13071

Martin M (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet.journal 17: 10–12. https://doi.org/10.14806/ej.17.1.200

Martins FMS, Porto M, Feio MJ, Egeter B, Bonin A, Serra SRQ, Taberlet P, Beja P (2020) Modelling technical and biological biases in macroinvertebrate community assessment from bulk preservative using multiple metabarcoding markers. Molecular Ecology. https://doi.org/10.1111/mec.15620

Meyer A, Boyer F, Valentini A, Bonin A, Ficetola GF, Beisel J-N, Bouquerel J, Wagner P, Gaboriaud C, Leese F, Dejean T, Taberlet P, Usseglio-Polatera P (2020) Morphological vs. DNA metabarcoding approaches for the evaluation of stream ecological status with benthic invertebrates: Testing different combinations of markers and strategies of data filtering. Molecular Ecology n/a. https://doi.org/10.1111/mec.15723

Mirtl M, T. Borer E, Djukic I, Forsius M, Haubold H, Hugo W, Jourdan J, Lindenmayer D, McDowell WH, Muraoka H, Orenstein DE, Pauw JC, Peterseil J, Shibata H, Wohner C, Yu X, Haase P (2018) Genesis, goals and achievements of Long-Term Ecological Research at the global scale: A critical review of ILTER and future directions. Science of The Total Environment 626: 1439–1462. https://doi.org/10.1016/j.scitotenv.2017.12.001

Nichols RV, Vollmers C, Newsom LA, Wang Y, Heintzman PD, Leighton M, Green RE, Shapiro B (2018) Minimizing polymerase biases in metabarcoding. Molecular Ecology Resources 18: 927–939. https://doi.org/10.1111/1755-0998.12895

Nielsen M, Gilbert MTP, Pape T, Bohmann K (2019) A simplified DNA extraction protocol for unsorted bulk arthropod samples that maintains exoskeletal integrity. Environmental DNA 1: 144–154. https://doi.org/10.1002/edn3.16

Pereira-da-Conceicoa L, Elbrecht V, Hall A, Briscoe A, Barber-James H, Price B (2019) Metabarcoding unsorted kick-samples facilitates macroinvertebrate-based biomonitoring with increased taxonomic resolution, while outperforming environmental DNA. bioRxiv: e792333. https://doi.org/10.1101/792333

Pont D, Valentini A, Rocle M, Maire A, Delaigue O, Jean P, Dejean T (2021) The future of fish-based ecological assessment of European rivers: from traditional EU Water Framework Directive compliant methods to eDNA metabarcoding-based approaches. Journal of Fish Biology 98: 354–366. https://doi.org/10.1111/jfb.14176

Ratnasingham S, Hebert PDN (2007) bold: The Barcode of Life Data System (http://www.barcodinglife.org). Molecular Ecology Notes 7: 355–364. https://doi.org/10.1111/j.1471-8286.2007.01678.x

Ruijter JM, Ramakers C, Hoogaars WMH, Karlen Y, Bakker O, Van den Hoff MJB, Moorman AFM (2009) Amplification efficiency: linking baseline and bias in the analysis of quantitative PCR data. Nucleic acids research 37: e45–e45. https://doi.org/10.1093/nar/gkp045

Schnell IB, Bohmann K, Gilbert MTP (2015) Tag jumps illuminated – reducing sequence-to-sample misidentifications in metabarcoding studies. Molecular Ecology Resources 15: 1289–1303. https://doi.org/10.1111/1755-0998.12402

Schrader C, Schielke A, Ellerbroek L, Johne R (2012) PCR inhibitors – occurrence, properties and removal. Journal of Applied Microbiology 113: 1014–1026. https://doi.org/10.1111/j.1365-2672.2012.05384.x

Sze MA, Schloss PD (2019) The Impact of DNA Polymerase and Number of Rounds of Amplification in PCR on 16S rRNA Gene Sequence Data. mSphere 4. https://doi.org/10.1128/mSphere.00163-19

Vamos EE, Elbrecht V, Leese F (2017) Short COI markers for freshwater macroinvertebrate metabarcoding. PeerJ Preprints. https://doi.org/10.7287/peerj.preprints.3037v2

Waskom M, Botvinnik O, Gelbart M, Ostblom J, Hobson P, Lukauskas S, Gemperline DC, Augspurger T, Halchenko Y, Warmenhoven J, Cole JB, De Ruiter J, Vanderplas J, Hoyer S, Pye C, Miles A, Swain C, Meyer K, Martin M, Bachant PQ, Kunter G, Villalba S, Brian, Fitzgerald C, Evans CG, Williams ML, O'Kane D, Yarkoni T, Brunner T (2020) seaborn: Statistical data visualization. Astrophysics Source Code Library: ascl:2012.015.

Yu DW, Ji Y, Emerson BC, Wang X, Ye C, Yang C, Ding Z (2012) Biodiversity soup: metabarcoding of arthropods for rapid biodiversity assessment and biomonitoring. Methods in Ecology and Evolution 3: 613–623. https://doi.org/10.1111/j.2041-210X.2012.00198.x

Zizka VMA, Leese F, Peinert B, Geiger MF (2018) DNA metabarcoding from sample fixative as a quick and voucher-preserving biodiversity assessment method1. Genome. https://doi.org/10.1139/gen-2018-0048

Zizka VMA, Elbrecht V, Macher J-N, Leese F (2019) Assessing the influence of sample tagging and library preparation on DNA metabarcoding. Molecular Ecology Resources 19: 893–899. https://doi.org/10.1111/1755-0998.13018

Supplementary material 1
Protocol 1 – DIY-DS
Authors: Dominik Buchner, Peter Haase, Florian Leese
Data type: text
Copyright notice: This dataset is made available under the Open Database License (http://opendatacommons.org/licenses/odbl/1.0/). The Open Database License (ODbL) is a license agreement intended to allow users to freely share, modify, and use this Dataset while maintaining this same freedom for others, provided that the original source and author(s) are credited.
Link: https://doi.org/10.3897/mbmg.5.67533.suppl1

Supplementary material 2
Table S1. PCR primers used in this study
Authors: Dominik Buchner, Peter Haase, Florian Leese
Data type: text
Copyright notice: This dataset is made available under the Open Database License (http://opendatacommons.org/licenses/odbl/1.0/). The Open Database License (ODbL) is a license agreement intended to allow users to freely share, modify, and use this Dataset while maintaining this same freedom for others, provided that the original source and author(s) are credited.
Link: https://doi.org/10.3897/mbmg.5.67533.suppl2

Supplementary material 3
Table S2. Raw read table
Authors: Dominik Buchner, Peter Haase, Florian Leese
Data type: text
Copyright notice: This dataset is made available under the Open Database License (http://opendatacommons.org/licenses/odbl/1.0/). The Open Database License (ODbL) is a license agreement intended to allow users to freely share, modify, and use this Dataset while maintaining this same freedom for others, provided that the original source and author(s) are credited.
Link: https://doi.org/10.3897/mbmg.5.67533.suppl3

Supplementary material 4
Table S3
Authors: Dominik Buchner, Peter Haase, Florian Leese
Data type: text
Copyright notice: This dataset is made available under the Open Database License (http://opendatacommons.org/licenses/odbl/1.0/). The Open Database License (ODbL) is a license agreement intended to allow users to freely share, modify, and use this Dataset while maintaining this same freedom for others, provided that the original source and author(s) are credited.
Link: https://doi.org/10.3897/mbmg.5.67533.suppl4

Supplementary material 5
Script 1
Authors: Dominik Buchner, Peter Haase, Florian Leese
Data type: code
Copyright notice: This dataset is made available under the Open Database License (http://opendatacommons.org/licenses/odbl/1.0/). The Open Database License (ODbL) is a license agreement intended to allow users to freely share, modify, and use this Dataset while maintaining this same freedom for others, provided that the original source and author(s) are credited.
Link: https://doi.org/10.3897/mbmg.5.67533.suppl5

Supplementary material 6
Figure S1. Pictures were taken with a digital microscope (Keyence VHX-6000, Keyence, Osaka, Japan)
Authors: Dominik Buchner, Peter Haase, Florian Leese
Data type: image
Explanation note: Top row: Kick-net samples. Bottom row: Malaise trap samples.
Copyright notice: This dataset is made available under the Open Database License (http://opendatacommons.org/licenses/odbl/1.0/). The Open Database License (ODbL) is a license agreement intended to allow users to freely share, modify, and use this Dataset while maintaining this same freedom for others, provided that the original source and author(s) are credited.
Link: https://doi.org/10.3897/mbmg.5.67533.suppl6

Supplementary material 7
Figure S2. Baseline-corrected amplification curves (left half) and melting -curves (right half) for A & B) kick-net samples and C & D) malaise trap samples
Authors: Dominik Buchner, Peter Haase, Florian Leese
Data type: statistical data
Explanation note: Different colors indicate different technical replicates of the same sample.
Copyright notice: This dataset is made available under the Open Database License (http://opendatacommons.org/licenses/odbl/1.0/). The Open Database License (ODbL) is a license agreement intended to allow users to freely share, modify, and use this Dataset while maintaining this same freedom for others, provided that the original source and author(s) are credited.
Link: https://doi.org/10.3897/mbmg.5.67533.suppl7

Supplementary material 8
Figure S3
Authors: Dominik Buchner, Peter Haase, Florian Leese
Data type: image
Copyright notice: This dataset is made available under the Open Database License (http://opendatacommons.org/licenses/odbl/1.0/). The Open Database License (ODbL) is a license agreement intended to allow users to freely share, modify, and use this Dataset while maintaining this same freedom for others, provided that the original source and author(s) are credited.
Link: https://doi.org/10.3897/mbmg.5.67533.suppl8